Sparse- vs. Dense Retrieval for Climate claim Verification
Design Document

Group 1
Dominik Kolak, Jonas Krauzer, Tristan Loidl

November 2025

1 Abstract

Climate change is a major threat to our society. While the 2025 United Nations Climate Change Conference did not yield satisfying results, and the 1.5-degree becomes increasingly uncertain, and fake news spreads over the internet. Because of that, we will build a Citation-Grounded Retrieval system for Fact-Checking, which allows the user to enter a claim and search for evidences that prove or falsify the given statement. This evidences are reranked and classified, returning one of the values *Supports*, *Refutes*, or *NotEnoughInformation* for the entered statement.

2 Idea

Our idea is build two Claim Verification Modules and evaluate how they compare in retrieving and classifying climate-related claims.

Approach 1   uses efficient sparse retrieval (BM25) to get the top k evidences, ranked by the BM25 score. The candidate's evidences are labeled using a BERT-based classifier, and concluding if a Claim is **True** or **False**.

Approach 2   uses more sophisticated sentence transformer on the same dataset. The top-k evidences with the highest dense vector similarity are then reranked using a more sophisticated cross-encoder and classified using the domain specific Climate-BERT classifier.

Dataset   We use the Climate-FEVER dataset as a data source. This dataset was created by the University of Zurich and contains 1535 real-world FEVER-style claims about climate change, each claim is manually annotated with five

evidence sentences that support, refute, or do not provide enough information to support or refute the claim.

`https://www.kaggle.com/datasets/bouweceunen/climate-fever-dataset` [30.11.2025]
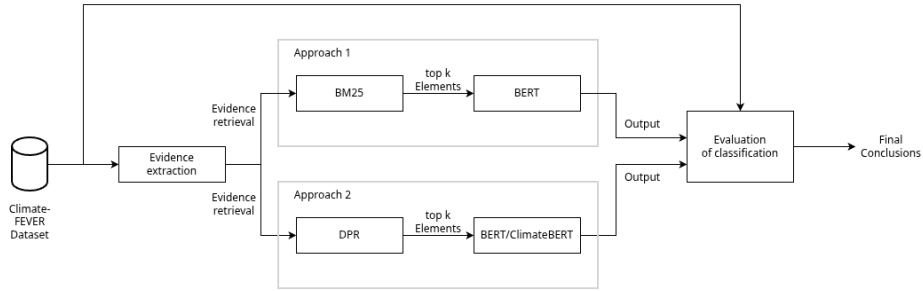
3 Visualization

Figure 1: Suggested approach

4 Evaluation

Both parts (retrieval and classification) are evaluated using different metrics. For the retrieval part, we will use Recall@k and Mean Reciprocal Rank (MRR), and for the classification, we will calculate the F1-Score and Accuracy. Calculation of accuracy should work, as the total size is not that high.

5 Tasks

- Extraction of evidences from the Climate-FEVER dataset

- Creation of sparse retrieval pipeline (Approach 1, simpler, faster to implement)

  - Implementation of the BM25 ranking algorithm
  - Implementation of the BERT cross-encoder

- Creation of dense retrieval pipeline (Approach 2, more complex, computationally more intensive, better semantic understanding expected)

  - Implementation of the DPR
  - Implementation of the Climate-BERT cross-encoder
  - Implementation of the Climate-BERT classifier

- Evaluation of the results by comparing both approaches with the original dataset
    - Evaluation of the retrieved documents
    - Evaluation of the Classification
- Discussion of results in the group, possibly finetuning and re-evaluation
- Preparation of the final presentation

Roles / Responsibilities

- Tristan Loidl: Approach 1
- Dominik Kolak: Approach 2
- Jonas Krauzer: Evaluation, Dataset preprocessing